Gene, 145 (1994) 163–169

GENE 08027

Non-cloning amplification of specific DNA fragments from whole genomic DNA digests using DNA 'indexers'

(Genome mapping; polymerase chain reaction; ligation; synthetic oligodeoxyribonucleotide; class-IIS restriction enzymes; STS mapping; RFLP; interrupted palindrome restriction enzymes)

Paul Unrau and Kenneth V. Deugau

AECL Research, Chalk River Laboratories, Chalk River, Ontario K0J 1J0, Canada

Received by F. Barany: 6 November 1993; Revised/Accepted: 7 March/10 March 1994; Received at publishers: 14 April 1994

SUMMARY

A highly systematic, non-cloning method of distinguishing and isolating every fragment in a class-IIS or interrupted palindrome restriction digest has been developed in our laboratory. These enzymes produce informative, non-identical cohesive ends which can be selectively modified by ligation to individual synthetic oligodeoxyribonucleotides with the corresponding complementary ends. In this way, polymerase chain reaction and sequencing primer sites and labels can be introduced specifically into a single fragment in a total genomic digest. Known and unknown fragments from genomes of the complexity of *Escherichia coli* can be isolated directly in sequenceable form without the necessity of synthesizing unique primers. Human DNA has also been accessed in this way. Problems intrinsic to cloning (selective fragment loss, mutation and sequence rearrangement) are avoided. Systematic characterization of DNA fragments by their cohesive ends and length provides tremendous power and flexibility for analysis of any DNA molecule without specific clones, probes or libraries. We report proof of principle of this remarkable system and indicate potential applications in DNA sequence tagged site and restriction mapping, sequencing, restriction-fragment-length polymorphism analysis and DNA diagnostics.

INTRODUCTION

Restriction fragments produced by class-IIS (Szybalski et al., 1991) and interrupted palindrome (IP) (Burger and Schinzel, 1983; Quiang and Schildkraut, 1984) restriction enzymes have non-identical end sequences. These informative ends can be ligated to synthetic DNA **indexers** to permit unique characterization of each fragment. Class-IIS enzymes cut a constant distance from their binding sites, and IP cut an internal DNA sequence which is not specified by the binding site (Fig. 1), so the overhanging sequences generated at any one cut site are invariant, but not predicted from the binding site sequence. Szybalski et al. (1991) reviewed some of the features and applications for class-IIS enzymes.

The specificity of the DNA ligase reaction is used to attach complementary **indexers** to DNA fragments. Under routine conditions ligation is specific for complementary cohesive ends (although forcing conditions can reduce specificity; Tsiapallis and Narang, 1970; Sgaramella and Khorana, 1972; Wiaderkiewicz and

Correspondence to: Dr. P. Unrau, Radiation Biology Branch, AECL Research, Chalk River Laboratories, Chalk River, Ontario, K0J 1J0, Canada. Tel: (1-613) 584-8811 ext. 6306; Fax: (1-613) 584-1713; e-mail: UNRAUP@CRL.AECL.CA

Abbreviations: bp, base pair(s); CDTA, *trans*-1,2-diaminocyclohexane-*N,N,N',N'*-tetraacetic acid (EDTA analog); CIAA, chloroform-isoamyl alcohol, 24:1; dNTP, deoxyribonucleoside triphosphate; EDTA, 1,2-diaminoethane-*N,N,N',N'*-tetraacetic acid; IP, interrupted palindrome(s); kb, kilobase(s) or 1000 bp; N, A or C or G or T; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; PCR, polymerase chain reaction; RFLP, restriction-fragment length polymorphism; STS, sequence tagged site(s); TC, 10 mM Tris·HCl pH 7.8/0.1 mM CDTA; TLC, thin-layer chromatography; TPE, 80 mM Tris·phosphate pH 7.8/8 mM EDTA; u, unit(s); YAC, yeast artificial chromosome.

SSDI 0378-1119(94)00238-N

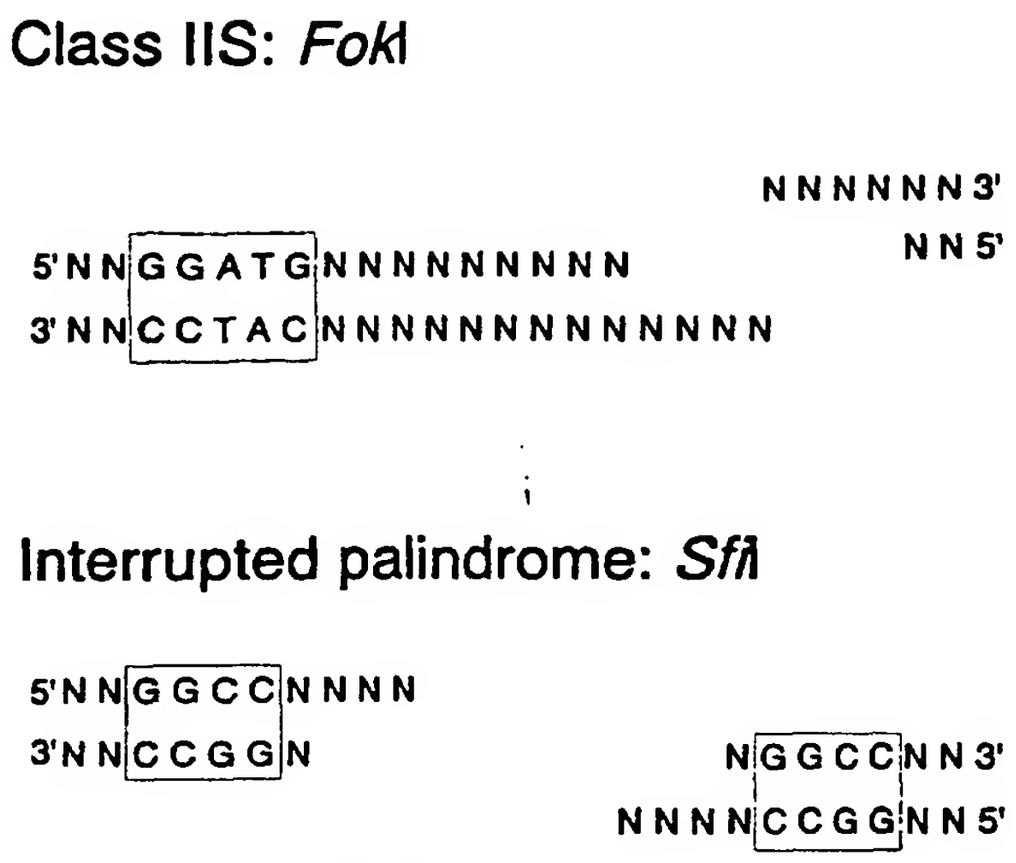

Fig. 1. Class-IIS and IP restriction endonuclease binding and cleavage sites. The class-IIS enzyme *Fok*I binds at the 5'-GGATG sequence and cuts 9 and 13 nt away to generate 4-nt 5'-cohesive ends. The IP enzyme *Sfi*I binds at the 5'-GGCCNNNNNGGCC sequence and cuts between the halves of the recognition sequence to generate 3-nt 3'-cohesive ends.

Ruiz-Carrillo, 198[illegible]We have used ligation to import selectively primer-binding sites or other structural features onto DNA target fragments. DNA fragments with known primer-binding sequences attached to both ends can be PCR amplified (Saiki et al., 1985; Kinzler and Vogelstein, 1989). Combinations of an internal, specific PCR primer and one external ligated adapter have also been used in this way (Patel et al., 1991; Smith, 1992).

When specific **indexers** have been attached to these end sequences, there is no necessity to know any part of the fragment sequence for PCR or sequencing. We demonstrate direct isolation of a variety of fragments from digests of λ, *E. coli* and human DNA by ligation of **indexers** followed by PCR. **Indexers** allow the selective definition and amplification of individual DNA fragments without cloning, opening up a powerful alternative route for analysis and sequencing of any DNA fragment.

TABLE I

Class-IIS and IP restriction enzymes which generate digestion fragments with informative cohesive ends

Enzyme	Recognition sequence[a]	End length[b]	Fragment length[c]	Fragment classes[d]
5'-Protruding cohesive ends:				
*Bbs*I	GAAGACN(2/6)	4	2024	32896
*Bbv*I	GCAGCN(8/12)	4	512	32896
*Bsa*I	GGTCTCN(1/5)	4	2024	32896
*Bsm*A1	GTCTCN(1/5)	4	512	32896
*Bsp*M1	ACCTGCN(4/8)	4	2024	32896
*Ear*I	CTCTTCN(1/4)	3	2024	2080
*Eco*31I	GGTCTCN(1/4)	3	2024	2080
*Esp*3I	CGTCTCN(1/5)	4	2024	32896
*Fok*I	GGATGN(9/13)	4	512	32896
*Hga*I	GACGCN(5/10)	5	512	524800
*Ksp*6321	CTCTCN(1/4)	3	512	2080
*Sap*I	GCTCTTCN(1/4)	3	8196	2080
*Sfa*NI	GCATCN(5/9)	4	512	32896
*Sts*I	GGATGN(10/14)	4	512	32896
3'-Protruding cohesive ends:				
*Apa*B15	GCANNNNN/TGC	5	4096	524800
*Bgl*I	GCCNNNN/NGGC	3	4096	2080
*Bst*X1	CCANNNNN/NTGG	4	4096	32896
*Dra*III	CACNNN/GTG	3	4096	2080
*Pfl*M1	CCANNNN/NTGG	3	4096	2080
*Rle*AI	CCCACAN(12/9)	3	2024	2080
*Sfi*I	GGCCNNNN/NGGCC	3	65536	2080

[a] The recognition sequence and position of cleavage site of each restriction enzyme are shown (Roberts and Macelis, 1992). N: any base; (*x*/*y*): number of nt between recognition site and cleavage site on each strand of the DNA.

[b] Length of the cohesive end (nt) generated by enzyme cleavage.

[c] Estimated assuming random base distribution, average fragment length (bp) will be 4^x, where x is the number of nt in the recognition sequence.

[d] The number of permutations of end sequences taken in pairs; for example, $(256 \times (255/2)) + 256 = 32896$ fragment classes are possible for fragments with 4-nt cohesive ends ($256 = 4^4$); the added 256 represent the classes of fragments with identical ends.

RESULTS AND DISCUSSION

(a) Class-IIS and IP restricti n enzymes

Class-IIS and IP enzyme digests are suitable for indexing because they produce non-identical end sequences. There are 4^x such end sequences, where x is the length of the overhang produced by digestion. These overhanging ends are the targets for **indexer** ligation. Table I lists class-IIS and IP enzymes of interest for indexing. Our initial studies have made use of two of these enzymes: *Fok*I is a class-IIS enzyme that cuts DNA every 512 bp, on average, and generates 4-nt 5′-cohesive ends; *Sfi*I is an IP enzyme that cuts rarely (about every 65 kb), generating 3-nt 3′-cohesive ends (Fig. 1). There are $4^4 = 256$ possible ends generated by *Fok*I digestion and $4^3 = 64$ possible ends generated by *Sfi*I digestion.

(b) Indexers

Indexers consist of an individually synthesized indexer strand annealed to a complementary common sequence (Fig. 2). Each **indexer** has a specific cohesive end sequence. The 64 different possible 3-nt ends generated by *Sfi*I require 64 **indexers** for complete access; the 256 different possible 4-nt *Fok*I-end sequences require 256 **indexers**.

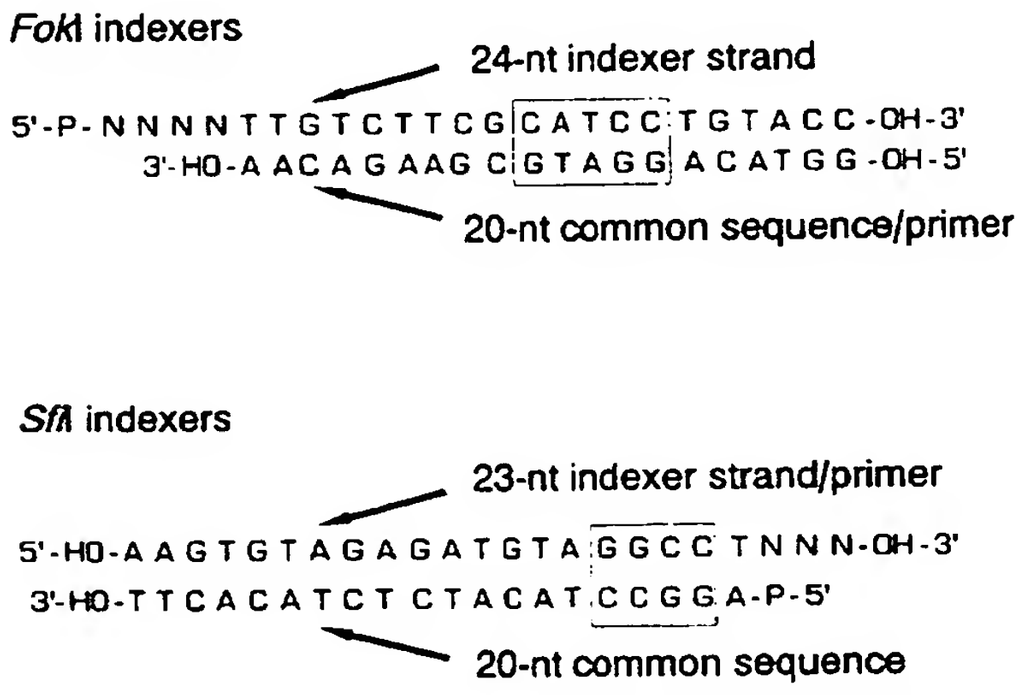

Fig. 2. **Indexer** structure. *Fok*I **indexers** have a 20-nt common sequence (#1504) annealed to one of 256 individually synthesized 24-nt complementary indexer strands with 4-nt 5′-cohesive ends. #1504 is the universal PCR primer for *Fok*I-indexed fragments. *Sfi*I **indexers** have a 20-nt common sequence (#3030-P) annealed to one of 64 individually synthesized 23-nt complementary indexer strands with 3-nt 3′-cohesive ends. The *Sfi*I-indexer strand is the PCR primer. Oligos were synthesized on a Milligen Biosearch DNA synthesizer and purified by TLC (Archer et al., 1985). *Fok*I-indexer strands and the *Sfi*I-common sequence were phosphorylated at the 5′-end. The two strands of the **indexers** were annealed by mixing equimolar quantities of an indexer strand with the appropriate common sequence at a concentration of 2 μM each in TC (10 mM Tris·HCl pH 7.8/0.1 mM CDTA pH 8.0), heating to 65 C for 5 min, and cooling at 1 C/10 s to room temperature using an MJ Research PTC-100. Annealed **indexers** were diluted to 50 nM in TC for use.

(c) Indexing known *Fok*I DNA fragments

To test indexing we ligated **indexers** to increasingly complex DNA digestions. The pBR322 plasmid has 12 *Fok*I-generated fragments (Watson, 1988). A 288-bp *Fok*I fragment with an internal *Pst*I site was identified between 3513 and 3800 bp in the pBR322 sequence. The 5′-overhanging ends on this *Fok*I fragment are 5′-CTTT and 5′-TAAG. **Indexers** with complementary ends (5′-pAAAG, 5′-pCTTA) were ligated to a pBR322 *Fok*I digest. The ligated DNA was amplified by PCR using the primer #1504, and the products analyzed by agarose-gel electrophoresis. There were no PCR products from unligated DNA or DNA digests that had been indexed with either the 5′-pAAAG or the 5′-pCTTA **indexer** alone. A PCR product was observed only when both **indexers** were present in the ligation. Its size was consistent with the predicted length of 288 bp plus 40 bp contributed by the **indexers**. Digestion of this PCR product with *Pst*I gave rise to the expected cleavage products (about 215 and 120 bp; Fig. 3).

Bacteriophage λ has 150 *Fok*I sites (Sanger et al., 1982). We synthesized a pair of **indexers** with 5′-pATAT and 5′-pAACG cohesive ends to target a 315-bp *Fok*I fragment from the λ genome (between 11 171 and 11 485 bp). The **indexers** were ligated to a λ *Fok*I digest, and the ligated DNA was amplified using the primer #1504. Again, unless both **indexers** were used in the ligation, no PCR products were observed. When both **indexers** were present, a PCR product of the predicted size (355 bp) was found. Digestion of this product with *Hha*I gave a cleavage pattern consistent with that predicted from the sequence (Fig. 4), and sequence analysis confirmed the identity of the fragment. This fragment was selected to demonstrate indexing with a self-complementary **indexer** end. Similar results confirming the specificity of indexing in genomes of increasing complexity have been obtained (A.C.Y. Chang and H.L. Heyneker, personal communication).

(d) Indexing unknown *Fok*I fragments by random access

There are about 9000 *Fok*I fragments in *E. coli*, assuming a random distribution of DNA sequences and a total genome of 4.67×10^6 bp (Perkins et al., 1992). In *E. coli* we predict that there are about 70 ligation targets for any given **indexer**; at random the other end of some of these target fragments will be complementary to other **indexers** being tested. Such fragments, indexed at both ends, will be amplification substrates. The complementary *Fok*I-end sequences of PCR products can be deduced from the **indexers** leading to fragment amplification. We estimate that the odds of obtaining a fragment with any two given ends in the whole *E. coli* genome is about 27%. The **indexers** synthesized to target pBR322 and λ *Fok*I frag-

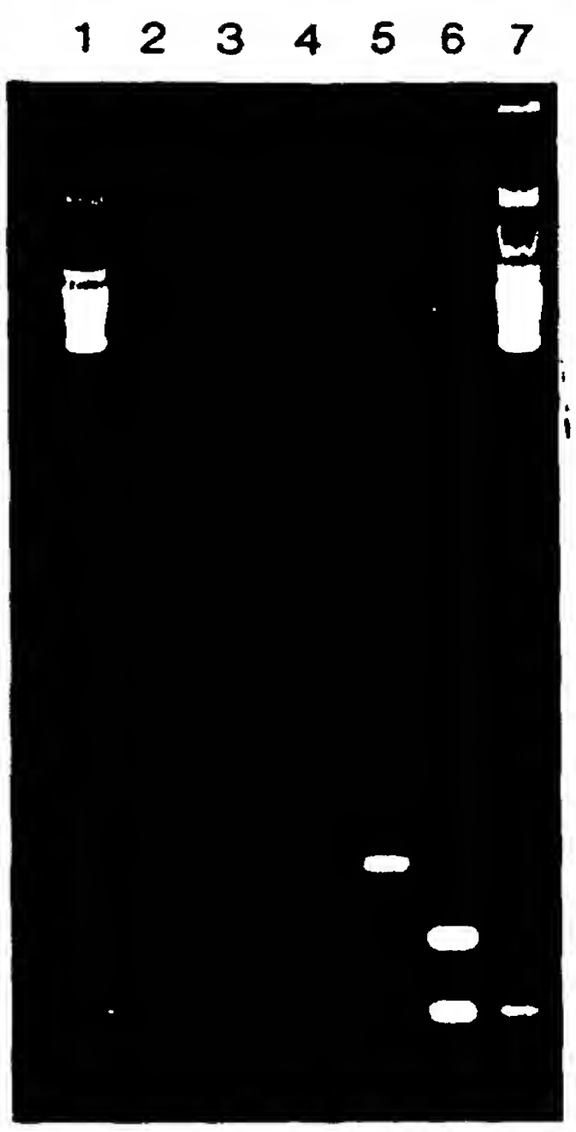

Fig. 3

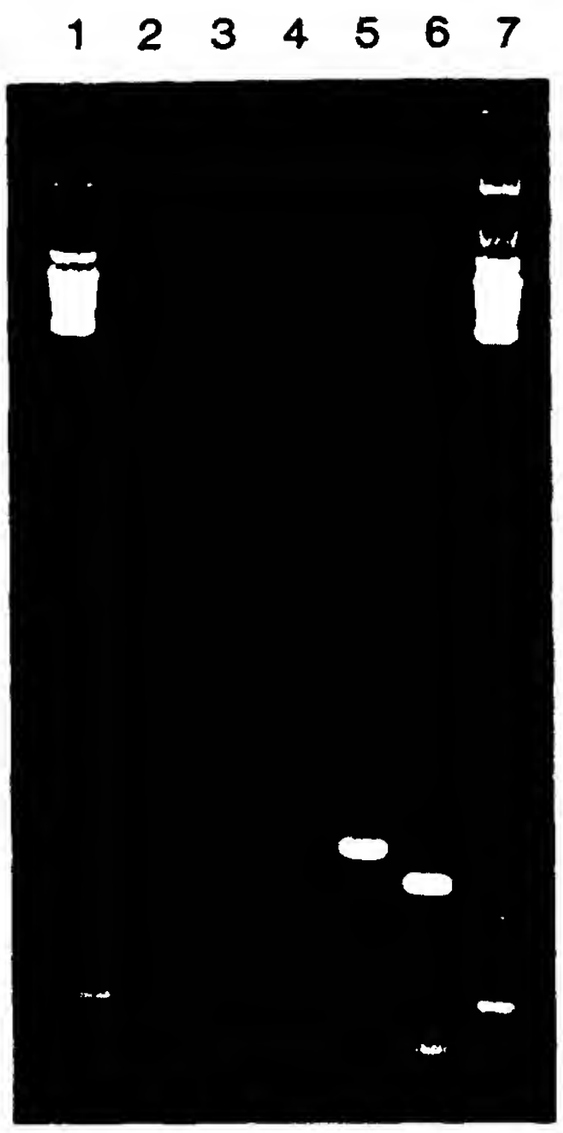

Fig. 4

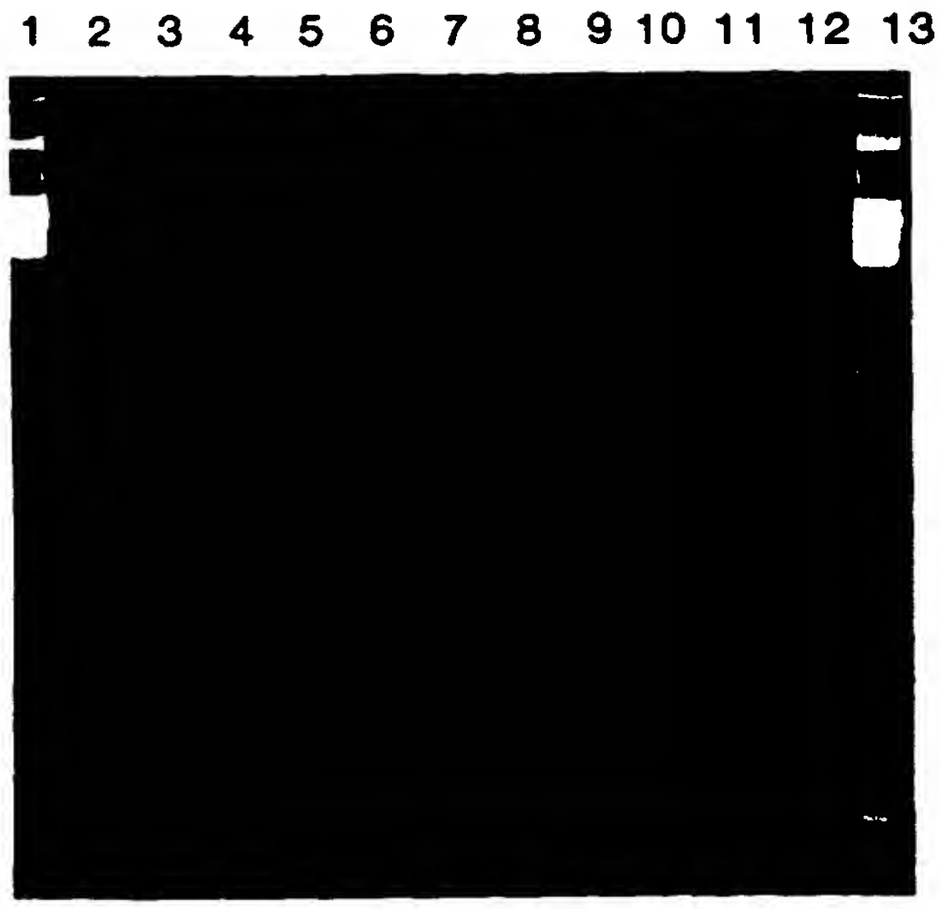

Fig. 5

Fig. 3. PCR amplification and enzyme digestion of an indexed pBR322 *Fok*I fragment. Lanes: **1** and **7**, 123-bp ladder DNA molecular weight standard (Gibco BRL, Life Technologies, Burlington, Ontario, Canada); **2**, no **indexers** in ligation (see **Methods** below); **3**, 5′-pAAAG **indexer** alone; **4**, 5′-pCTTA **indexer** alone; **5**, both **indexers** together giving a 330-bp PCR product; **6**, digestion of the product with *Pst*I (Gibco BRL) giving 215- and 120-bp fragments. **Methods:** pBR322 DNA (NE Biolabs, Beverly, MA, USA) was digested with *Fok*I (NE Biolabs) at 2 u/μg DNA for 30 min at 20°C, extracted with 1/2 vol. CIAA (24:1), and ethanol precipitated. DNA was recovered by microcentrifugation at 14000 × *g* for 30 min, washed with 70% ethanol, air-dried, and dissolved in TC to 200 ng/μl. DNA (100 ng) was ligated for 30 min at 37°C in reactions containing 4 u T4 DNA ligase (NE Biolabs) and 50 fmol of each **indexer** (as specified for each lane) in a total volume of 20 μl recommended buffer. Reactions were terminated by heating to 95°C for 5 min. 10 ng of DNA from each ligation were amplified by PCR in an MJ Research PTC-100 using the following program: (1) 95°C, 5 min; (2) 80°C, hold during addition of 2 u Ampli*Taq* DNA polymerase (Perkin Elmer, Nepean, Ontario, Canada); (3) 94°C, 1 min 15 s; (4) 55°C, 30 s; (5) 72°C, 1 min 30 s; (6) cycle to step 3, 29 times; (7) 72°C, 5 min; (8) 15°C, hold. PCR reactions contained 1 mM $MgCl_2$ and 40 pmol primer #1504 in 50 μl of the recommended buffer, covered by 50 μl light mineral oil (Sigma, St. Louis, MO, USA). PCR reactions containing DNA fragments were extracted with 100 μl CIAA, ethanol-precipitated and resuspended in TC. A sample was then digested with 5 u *Pst*I at 37°C for 1 h in the recommended buffer. Amplification products were electrophoresed at 60 V for 1 h on 1.8% agarose gels in TPE buffer containing 500 ng/ml ethidium bromide.

Fig. 4. PCR amplification and enzyme digestion of an indexed λ *Fok*I fragment. Lanes: **1** and **7**, 123-bp ladder DNA molecular weight standard; **2**, no **indexers** in ligation; **3**, 5′-pATAT **indexer** alone; **4**, 5′-pAACG **indexer** alone; **5**, both **indexers** together giving a 355-bp PCR product; **6**, fragment digested with *Hha*I (NE Biolabs) giving 295- and 60-bp fragments. λ DNA (NE Biolabs) was digested with *Fok*I, ligated to **indexers**, and PCR amplified as described in Fig. 3. PCR reactions containing DNA fragments were digested with 7.5 u *Hha*I at 37°C for 1 h in the recommended buffer and electrophoresed as described in Fig. 3.

Fig. 5. PCR amplification of indexed *Fok*I fragments from whole genomic digests of *E. coli* B. Lanes: **1** and **13**, 123-bp ladder DNA molecular weight standard; **2**, no **indexers**; **3**, 5′-pAAAG **indexer** alone; **4**, 5′-pCTTA **indexer** alone; **5**, 5′-pATAT **indexer** alone; **6**, 5′-pAACG **indexer** alone; **7**, 5′-pAAAG + 5′-pCTTA **indexers**; **8**, 5′-pAAAG + 5′-pATAT **indexers**; **9**, 5′-pAAAG + 5′-pAACG **indexers**; **10**, 5′-pCTTA + 5′-pATAT **indexers**; **11**, 5′-pCTTA + 5′-pAACG **indexers**; **12**, 5′-pATAT + 5′-pAACG **indexers**. *E. coli* B DNA (Sigma) was digested with *Fok*I, ligated to **indexers**, and PCR amplified and electrophoresed as described in Fig. 3.

ments were ligated to *Fok*I-digested *E. coli* B DNA to see if fragments could be indexed and amplified from such a complex fragment pool. Fig. 5 shows the results of amplifying *E. coli* B *Fok*I-digested DNA indexed with the ten single and pair-wise combinations of the four **indexers**. Two unique fragments were obtained, present only when both the 5′-pCTTA and 5′-pATAT **indexers** are used. The fragment-end sequences must, therefore, be 5′-TAAG and 5′-ATAT for both fragments. These results are evidence that there is a good chance of indexing a fragment entirely at random in a whole genomic digest of *E. coli* with any given **indexer** or pair of **indexers**.

Recall that the two cohesive-end sequences of a DNA fragment are independent. Using both end sequences, 32896 distinct sets of *Fok*I fragments can be defined. In *E. coli* most pairs of **indexers** will not index an amplifiable

DNA fragment, as only 9000 fragmen expected vs. 32 896 unique pairs of **ind x rs**. Poisson statistics predict that about 10% of the defined subsets will contain more than one fragment. These sets of fragments with the same end sequences can be separated by length.

Under the conditions of ligation used, the extent of joining of **indexers** to any of the twelve possible single bp mismatched ends is not apparent. Even a limited amount of such 'misligated' product would be amplifiable by PCR, resulting in multiple product bands for every pair of **indexers**. Preliminary experiments to measure the rate of ligation at mispaired ends indicated that this rate is less than 10^{-3} of the correct ligation rate (unpublished data).

(e) Indexing DNA fragments with dissimilar cohesive ends

To see if different types of cohesive-end sequences can be accessed simultaneously, we indexed *Fok*I-*Sfi*I fragments in the SV40 genome. SV40 contains a single *Sfi*I site (Fiers et al., 1978). The cohesive ends generated by *Sfi*I were identified, as were the cohesive ends at the flanking *Fok*I sites (Fig. 6). Complementary **indexers** were synthesized to access the individual Left hand and Right hand fragments at this *Sfi*I site. Fig. 7 shows the results of indexing and amplifying the SV40 *Fok*I-*Sfi*I fragments. Amplification of the target fragments required both the *Fok*I and *Sfi*I primers; amplification with either primer alone did not give rise to a PCR product.

We have proposed using indexed fragments centered on *Sfi*I sites in STS mapping (Unrau et al., 1990). The *plsB* gene in *E. coli* contains an *Sfi*I site with adjacent *Fok*I sites located 1436 bp and 472 bp away. Fig. 8 shows the indexed amplification of these two fragments. Sequence analysis of the smaller fragment confirmed its identity as the predicted fragment of the *plsB* gene. This result also demonstrated that sequenceable templates can be generated directly from genomic DNA by indexed amplification.

(f) Indexing human genomic DNA

We have tested DNA indexing in human genomic DNA. The frequency of cleavage of human DNA by *Not*I and *Sfi*I suggested that a number of *Not*I-*Sfi*I fragments would be short enough to permit PCR amplification. Using one **indexer** complementary to the *Not*I end (5′-HO-GGCC) and specific **indexers** complementary to possible *Sfi*I ends we amplified six fragments (Fig. 9). Preliminary sequence data suggests that these products are specific and further characterization is in progress to confirm their identity. Religation of *Not*I ends and amplification of the few short *Not*I-*Not*I fragments were prevented by using an excess of the 5′-OH **indexer**.

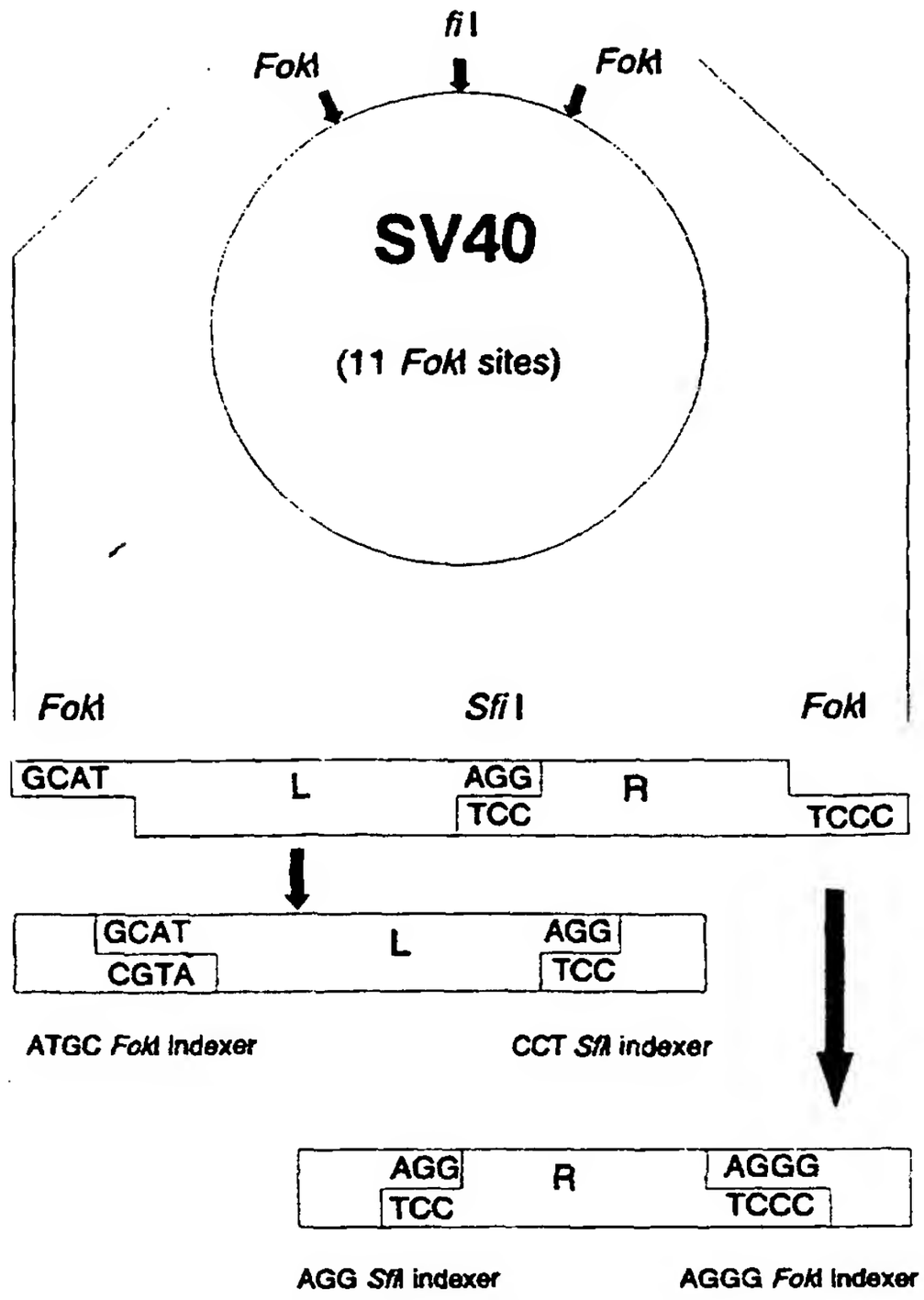

Fig. 6. SV40 *Fok*I-*Sfi*I sites targeted for indexing. *Sfi*I cuts the circular SV40 genome at a single site (5238 bp, Fiers et al., 1978) producing AGG-3′ and CCT-3′ cohesive ends. Cleavage of the adjacent *Fok*I sites (at 5049 bp and 97 bp) generates 5′-pCCCT and 5′-pGCAT cohesive ends. The two *Fok*I-*Sfi*I fragments can be indexed using CCT-3′ and 5′-pATGC **indexers**, and AGG-3′ and 5′-pAGGG **indexers**, shown joined to the target fragments by ligation. L and R refer to the left and right sides of the *Sfi*I site, respectively.

(g) Conclusions

(*1*) We describe a non-cloning method for defining, amplifying and isolating specific DNA fragments from complex genomic digests. This method depends upon the selectivity of ligation to attach **indexers** to unknown DNA fragments, importing primer-binding sites or other structural features which may be useful in fragment manipulation, including reporter or affinity groups.

(*2*) The biological variables which affect cloning do not affect indexing technology.

(*3*) The data demonstrate the specific, selective ligation of **indexers** to complementary end sequences in DNA digests. **Indexers** import a primer-binding site to the non-identical cohesive-end sequences of fragments produced by class-IIS and IP enzyme digestion. On PCR, such indexed fragments give rise to amplification products that

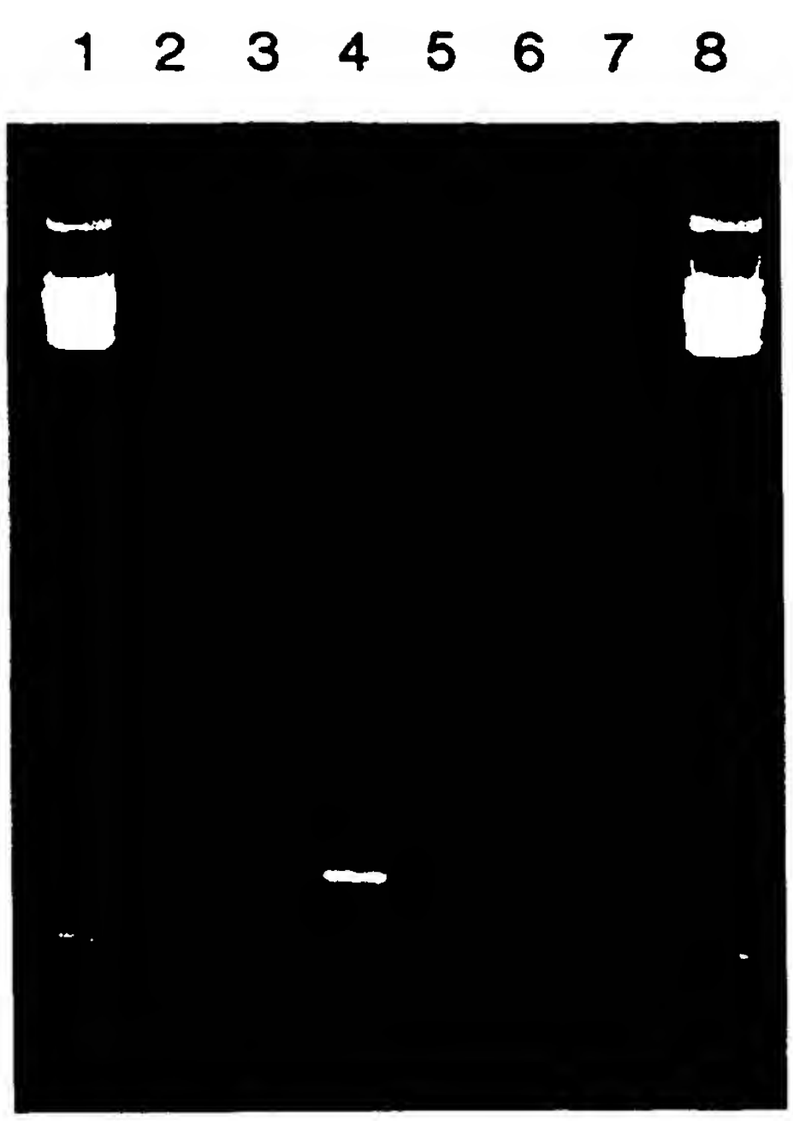

Fig. 7

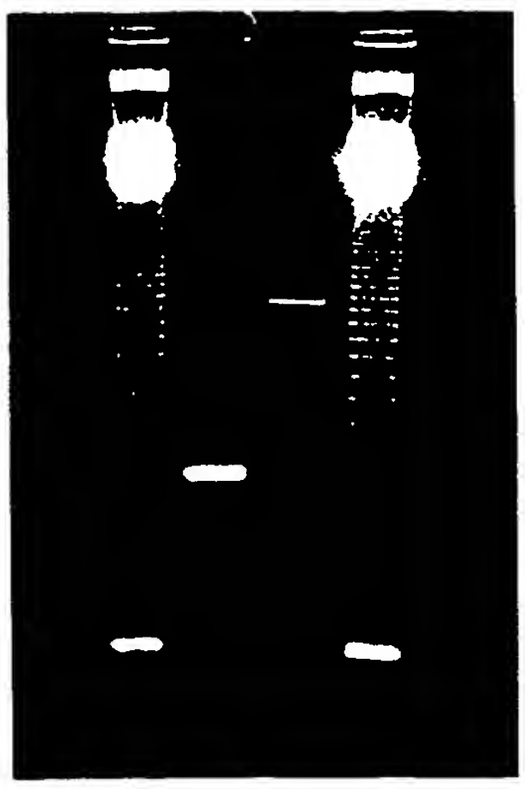

Fig. 8

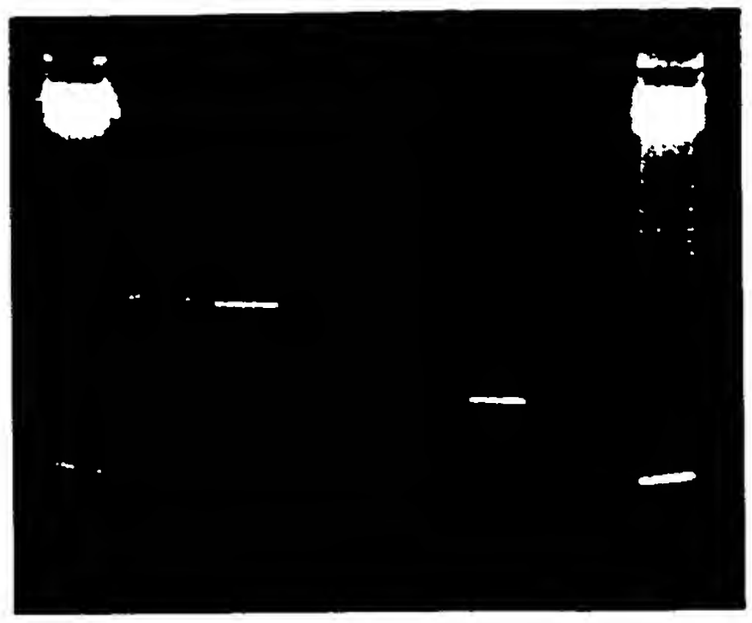

Fig. 9

Fig. 7. PCR amplification of indexed *Fok*I-*Sfi*I fragments of SV40. Lanes: **1** and **8**, 123-bp ladder DNA molecular weight standard; **2**, ligation (*i*) amplified with PCR primer #1504; **3**, ligation (*i*) amplified with PCR primer CCT-3′ *Sfi*I-indexer strand; **4**, ligation (*i*) amplified with PCR primers #1504 and CCT-3′ *Sfi*I-indexer strand; **5**, ligation (*ii*) amplified with PCR primer #1504; **6**, ligation (*ii*) amplified with PCR primer AGG-3′ *Sfi*I-indexer strand; **7**, ligation (*ii*) amplified with PCR primers #1504 and AGG-3′ *Sfi*I-indexer strand. **Methods:** SV40 DNA (Gibco BRL) was digested with *Fok*I (see Fig. 3), followed by two digestions with 2 u *Sfi*I (NE Biolabs)/μg DNA at 50°C overnight under oil. The digest was extracted with CIAA, ethanol precipitated and dissolved to 200 ng/μl in TC. Conditions for ligations and PCR amplification were as described in Fig. 3. Ligations contained 50 fmol of the following **indexers**: (*i*) 5′-pATGC *Fok*I **indexer** and CCT-3′ *Sfi*I **indexer**; (*ii*) 5′-pAGGG *Fok*I **indexer** and AGG-3′ *Sfi*I **indexer**. PCR reactions were analyzed by electrophoresis as described in Fig. 3.

Fig. 8. Amplification of indexed *E. coli plsB* gene (GenBank accession No. K00127) *Fok*I-*Sfi*I fragments. Lanes: **1** and **4**, 123-bp ladder DNA molecular weight standard; **2**, ligation (*i*) amplified with PCR primer #1504 and PCR primer GGC-3′ *Sfi*I-indexer strand; **3**, ligation (*ii*) amplified with PCR primers #1504 and GCC-3′ *Sfi*I-indexer strand. *E. coli* B DNA (Sigma) was digested with *Fok*I (see Fig. 3) followed by digestion twice with *Sfi*I (as described in Fig. 7). The digest was extracted with CIAA, ethanol precipitated and dissolved to 200 ng/μl in TC. Conditions for ligations and PCR amplification were as described in Fig. 3. Ligations contained 50 fmol of the following **indexers**: (*i*) 5′-pTCCG *Fok*I **indexer** and GGC-3′ *Sfi*I **indexer**; (*ii*) 5′-pCGCG *Fok*I **indexer** and GCC-3′ *Sfi*I **indexer**. PCR reactions were analyzed by electrophoresis as described in Fig. 3.

Fig. 9. Amplification of indexed human *Not*I-*Sfi*I fragments. Lanes: **1** and **8**, 123-bp ladder DNA molecular weight standard; **2**, ligation (*i*) amplified with primer #1504 and CCA-3′ *Sfi*I-indexer strand; **3**, ligation (*ii*) amplified with primer #1504 and CCT-3′ *Sfi*I-indexer strand; **4**, ligation (*iii*) amplified with primer #1504 and TAC-3′ *Sfi*I-indexer strand; **5**, ligation (*iv*) amplified with primer #1504 and TAG-3′ *Sfi*I-indexer strand; **6**, ligation (*v*) amplified with primer #1504 and CGG-3′ *Sfi*I-indexer strand; **7**, ligation (*vi*) amplified with primer #1504 and CTG-3′ *Sfi*I-indexer strand. Human DNA [isolated from cultured LCL cells essentially as described by Jeanpierre (1987)] was twice digested with *Sfi*I as indicated in Fig. 7 and with *Not*I (NE Biolabs) 2 u/μg for 2 h at 37°C and recovered as indicated in Fig. 3. Ligation and PCR conditions (40 cycles) are as described in Fig. 3. Ligations contained 100 ng digested DNA, 50 fmol of the 5′-HO-GGCC *Fok*I **indexer** (complementary to the *Not*I end) and 50 fmol of one of the following *Sfi*I **indexers**: (*i*) CCA-3′; (*ii*) CCT-3′; (*iii*) TAC-3′; (*iv*) TAG-3′; (*v*) CGG-3′; (*vi*) CTG-3′. PCR reactions were analyzed by electrophoresis as described in Fig. 3.

correspond in length and restriction enzyme cleavage sites to those predicted from whole sequence analysis.

(*4*) Non-specific amplification of artefacts was not observed with the primers used.

(*5*) Ligation reactions can attach 4-nt 5′-protruding, or 3-nt 3′-protruding **indexers** simultaneously to fragments with different sorts of ends, including self-complementary ends.

(*6*) A single primer suffices for PCR of indexed *Fok*I fragments, while both *Fok*I and *Sfi*I primers are required for PCR of fragments indexed with both *Fok*I and *Sfi*I **indexers**. This proves that we are specifically importing primer-binding sites to indexed fragments.

(*7*) In whole genomic digests of *E. coli* B a limited set of fragments amplify after ligation to all pairwise combinations of four **indexers**. Successful indexing of *Fok*I digests of *E. coli* genomic DNA shows that selective ligation has the specificity to subdivide complex mixtures of DNA fragments in an orderly way.

(*8*) Because the number of possible cohesive ends is

limited, this method could be systematically applied both to fragment isolation and to access of information about each fragment.

(9) Because DNA fragments can be directly indexed, amplified and sequenced, it is reasonable to expect that indexing could be applied to the mapping and sequencing of YACs (Burke et al., 1987), isolated yeast chromosomes, whole bacterial genomes, and human genomic DNA. Indexing applied to detection of mutations at the cohesive ends of fragments would result in a variant of RFLP analysis called Restriction-Fragment End Polymorphisms (RFEPs) potentially useful for DNA diagnostic studies.

ACKNOWLEDGEMENTS

We acknowledge the support and interest of AECL Research in funding this work and of N.E. Gentner and D.P. Morrison for support and helpful discussions. We thank S. Laffrenier for skilled technical assistance in repeating the original indexing experiments for publication, and P. Burchart, D. Cecil and S. Adams for their skilled technical assistance in preparing and indexing the human DNA. The frank discussions, editorial revisions and comments of S. Carlisle are deeply appreciated. Sequence analysis of the λ fragment was confirmed in the laboratory of H.L. Heyneker by A.C.Y. Chang and the sequence of the *E. coli plsB* fragment was confirmed by K. McDermott and T. Inalsingh of the Queen's University Core Facility with our thanks. Financial support from CANDU Owners Group is gratefully acknowledged.

REFERENCES

Archer, T.K., Tam, S.-P., Deugau, K.V. and Deeley, R.G.: Apolipoprotein CII mRNA levels in primate liver: induction by estrogen in the human hepatocarcinoma cell line. J. Biol. Chem. 260 (1985) 1670–1675.

Burger, K.J. and Schinzel, R.: Restriction endonuclease *BglI* as a tool for in vivo reconstruction and recombination of plasmid and bacterial genomes. Mol. Gen. Genet. 189 (1983) 269–274.

Burke, D.T., Carle, G.F. and Olson, M.V.: Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236 (1987) 806–812.

Fiers, W., Conteras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volkaert, G. and Ysebaert, M.: Complete nucleotide sequence of SV40 DNA. Nature 273 (1978) 113–119.

Jeanpierre, M.: A rapid method for the purification of DNA from blood. Nucleic Acids Res. 15 (1987) 9611.

Kinzler, K.W. and Vogelstein, B.: Whole genome PCR: application to the amplification of sequences bounded by regulatory proteins. Nucleic Acids Res. 17 (1989) 3645–3653.

Patel, K.R., Cox, J., Shipley, F., Kiley, K., Fraser, D.R., Cox, H., Lebrach, H. and Sheer, D.: A novel and rapid method for isolating sequences adjacent to rare cutting sites and their use in physical mapping. Nucleic Acids Res. 19 (1991) 4371–4375.

Perkins, J.D., Heath, J.D., Sharma, B.R. and Weinstock, G.M.: *SfiI* genomic cleavage map of *Escherichia coli* K-12 strain MG1655. Nucleic Acids Res. 20 (1992) 1129–1137.

Quiang, B.-Q. and Schildkraut, I.: A type II restriction endonuclease with an eight base nucleotide specificity from *Streptomyces fimbriatus*. Nucleic Acids Res. 12 (1984) 4507–4515.

Roberts, R.J. and Macelis, D.: Restriction enzymes and their isoschizomers. Nucleic Acids Res. 20 (1992) 2167–2180.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N.: Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230 (1985) 1350–1354.

Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F. and Peterson, G.B.: Nucleotide sequence of bacteriophage (lambda) DNA. J. Mol. Biol. 162 (1982) 729–773.

Sgaramella, V. and Khorana, H.G.: CXII. Total synthesis of the structural gene for an alanine transfer RNA from yeast. Enzymatic joining of the chemically synthesized polydeoxynucleotides to form the DNA duplex representing nucleotide sequence 1–20. J. Mol. Biol. 72 (1972) 427–444.

Smith, D.R.: Ligation mediated PCR of restriction fragments from large DNA molecules. PCR Methods Appl. 2 (1992) 21–25.

Szybalski, W., Kim, S.C., Hasan, N. and Podhajska, A.J.: Class-IIS restriction enzymes – a review. Gene 100 (1991) 13–26.

Tsiapallis, C.M. and Narang, S.A.: On the fidelity of phage T4-induced polynucleotide ligase in the joining of chemically synthesized deoxyribonucleotides. Biochem. Biophys. Res. Commun. 39 (1970) 631–636.

Unrau, P. and Deugau, K.V.: A simple comprehensive method for functional definition, isolation and characterization of STS for genome mapping. Human Genome II Abst. No. 134 (1990).

Watson, N.: A new revision of the sequence of plasmid pBR322. Gene 70 (1988) 399–403.

Wiaderkiewicz, R. and Ruiz-Carrillo, A.: Mismatch and blunt to protruding-end joining by DNA ligase. Nucleic Acids Res. 15 (1987) 7831–7848.